Statistiques, cours, première S

1 Quantiles d'une série statistique

Définition :

- Le *premier quartile* noté Q_1 de la série statistique est ....
.... ;
- Le *troisième quartile* noté Q_3 de la série est ....
....

Détermination pratique :

On suppose la série ordonnée dans l'ordre croissant des valeurs du caractère. Soit N l'effectif total.

- Si $\frac{N}{4}$ est une entier alors Q_1 est la valeur de rang .... et Q_3 est la valeur de rang .... ;
- si $\frac{N}{4}$ n'est pas entier, alors Q_1 est la valeur dont le rang suit .... et Q_3 la valeur dont le rang suit ....

Remarque :

Il existe d'autres définitions des quartiles. En particulier, on peut définir le premier quartile comme la médiane de la sous série formée des valeurs inférieures ou égales à la médiane et le troisième quartile comme la médiane de la sous série formée des valeurs supérieures ou égales à la médiane.

Définition :

Le n^e *décile* est la plus petite valeur telle que .... des effectifs au moins lui soient inférieurs ou égales.

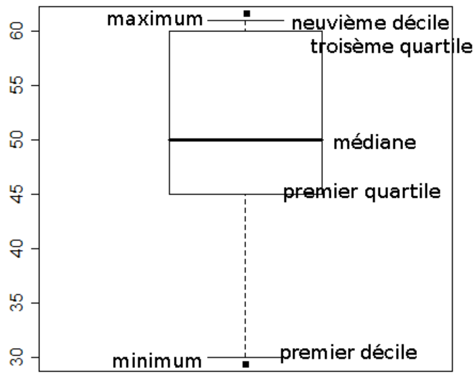
Définition :

Un *diagramme* en boîte représente les caractéristiques statistiques suivantes :

- .... ;
- .... ;
- .... ;
- ....

Remarque :

On fait parfois figurer le minimum et le maximum de la série au lieu des premier et neuvième déciles.



2 Caractéristiques de dispersion

On considère une série statistique $(n_i; x_i)_{i=1}^{p}$ d'effectif total $N = \sum_{i=1}^{p} n_i$.

Définition :

La *variance* de la série est le nombre noté V défini par :

....

Propriété :

....

Preuve :

$V = ...$

$V = ...$

$V = ...$

$V = ...$

$V = ...$

$V = ...$

$V = ...$

Propriété :

Pour tout réel x, on a :

$$\sum_{i=1}^{p} n_i(x_i - x)^2 \geq \sum_{i=1}^{p} n_i(x_i - \bar{x})^2$$

La variance minimise donc la distance des valeurs du caractère à la moyenne, c'est donc un indicateur approprié pour mesurer la dispersion des valeurs autour de la moyenne.

Preuve :

Soit f définie par $f(x) = \sum_{i=1}^{p} n_i(x_i - x)^2$ pour tout réel x. f est dérivable sur $\mathbb{R}$.

On a $f'(x) = -\sum_{i=1}^{p} 2n_i(x_i - x) = -2\sum_{i=1}^{p} n_i x_i + 2x\sum_{i=1}^{p} x_i = -2\sum_{i=1}^{p} n_i x_i + 2Nx$.

Donc $f'(x) \leq 0 \Leftrightarrow -\sum_{i=1}^{p} n_i x_i + aN \leq 0 \Leftrightarrow x \leq \frac{1}{N}\sum_{i=1}^{p} n_i x_i \Leftrightarrow x \leq \bar{x}$. Donc f est décroissante sur $]-\infty; \bar{x}]$ et croissante sur $[\bar{x}; +\infty[$ donc minimale pou $x = \bar{x}$.

Définition :

L'écart-type de la série statistique est le nombre noté σ défini par :

....

Définition :

On appelle *intervalle inter quartile* l'intervalle .... et *écart inter quartile* le nombre ....

3 Transformation affine de valeurs

Propriété :

Soient a et b deux réels et x_i pour i allant de 1 à p le valeurs non nécessairement distinctes d'une série statistique.
On note $\bar{x}$ sa moyenne, s son écart type, V sa variance, M_e, Q_1 et Q_3 sa médiane et ses quartiles.
On définit alors une nouvelle série en posant $X_i = ax_i + b$.

La moyenne de la nouvelle série est alors ...., sa variance ...., son écart type .... et, si $a > 0$, son premier quartile est ...., sa médiane ...., son troisième quartile .... et son écart inter quartiles est ....

Preuve :

$$\begin{aligned} \overline{aX_i + b} &= \ldots\ldots\ldots\ldots\ldots\ldots \\ &= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &= \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &= \ldots\ldots\ldots\ldots \end{aligned}$$

D'autre part,

$$\begin{aligned} \frac{\sum_{i=1}^{p}(X_i - \bar{X})^2}{N} &= \ldots\ldots\ldots\ldots \\ &= \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &= \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &= \ldots\ldots\ldots \end{aligned}$$

D'où l'écart type de la série des X_i est donnée par $\sqrt{a^2V} = |a|\sqrt{V} = |a|\,s$.

Les propriétés de la médiane et des quartiles découlent directement de la définition et de la conservation de l'ordre puisque $a > 0$.